Journal of Engineering Mathematics **43:** 105–114, 2002.

Mathematical modeling of polish-rate decay in chemical-mechanical polishing

L. BORUCKI
Motorola, Inc., 2200 W. Broadway Rd, M360, Mesa, Arizona, 85202, U.S.A. (E-mail: Len.Borucki@Motorola.com)

Received 11 June 2001; accepted in revised form 3 April 2002

Abstract. A model is presented for polish-rate decay in chemical-mechanical polishing based on the Greenwood-Williamson theory of contact between a smooth surface (a wafer) and a rough surface (the polishing pad). The model assumes that polishing causes pad asperities to wear, with high asperities wearing faster than low asperities. Model predictions of the time dependence of polish-rate decay compare favorably with experiments.

Key words: chemical-mechanical polishing, Hamilton-Jacobi equation, polymer wear, rate decay, surface statistics

1. Introduction

Depth-of-focus limitations of advanced lithography processes in the semiconductor industry require that surface layers that are grown or deposited on silicon wafers maintain a high degree of overall surface planarity. Many techniques have been developed for this purpose, but most do not meet the stringent requirements of the most advanced, multilevel metallization processes. The current preferred method for advanced planarization is Chemical-Mechanical Polishing (CMP), which has been shown in many cases to yield the best results for wafer planarity and uniformity.

In one large class of single-wafer CMP tools, the wafer is held upside down by a rotating wafer carrier while being pressed against a soft rotating closed pore polyurethane polishing pad or pad stack as shown in Figure 1. A chemically-reactive slurry containing a small weight fraction of fine (100 nm) abrasive particles is sprayed on the pad ahead of the wafer. Usually, a diamond-covered rotating disk called a conditioner is also swept back and forth radially across the pad during wafer polishing. The purpose of the conditioner is to refresh and roughen the pad surface. Pad-surface asperities that are tall enough to touch the wafer trap abrasive particles and drag them across the surface. This abrasive action, combined with chemical attack of the exposed wafer surface by the slurry, is thought to be responsible for polishing. Although the broad outlines of the mechanisms underlying the CMP process are qualitatively understood, many aspects of the process are not understood in detail and have not been modeled.

One of the areas that is not well understood is the connection between conditioning and the behavior of the average polish rate. A freshly conditioned pad usually produces a high polish rate. However, if the pad is not conditioned, the polish rate drops rapidly with time. A decrease of a factor of two or more in the rate is possible after an hour of unconditoned pad

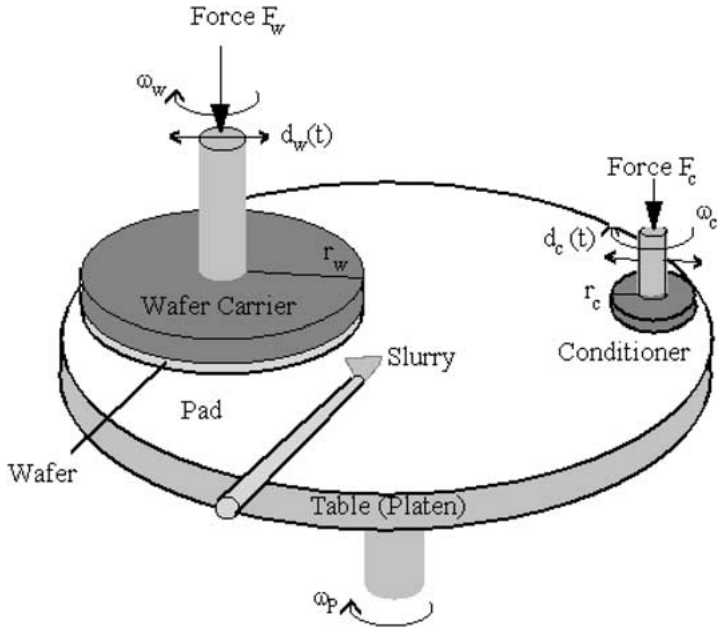

Figure 1. Schematic of a typical single-wafer CMP tool.

use. This is very undesirable from a manufacturing perspective because it affects throughput and cost.

The cause of polish-rate decay has been somewhat of a mystery. One polish-rate study [1], which used optical interferometry to characterize pad surfaces automatically, found little or no visible change in pad appearance or surface-roughess statistics (standard deviation, skewness, kurtosis) after the first few minutes of polishing of a thick layer of deposited silicon dioxide, even though the polish rate declined continuously. This suggested that polish-rate decay is not caused by changes in pad-surface morphology. However, another study by a pad manufacturer [2], also using interferometry but with filtering of points far from the surface plane of the pad, detected wear of interpore asperities that was correlated with a decrease in oxide-removal rate, suggesting that there is a connection between rate decay and surface morphology.

Here, we present a simple model of polish-rate decay that explains the polish-rate time dependence seen in the measurements of [1], while at the same time explaining why significant changes in the overall pad-surface statistics were not evident. The model at the same time is in agreement with the observations of [2].

2. Polishing-pad physical properties

Some discussion of CMP pad physical properties will be required for understanding of the polish-rate-decay model. We focus on the Rodel Corp. IC1000 pad, which is often used for planarizing silicon-dioxide films. The IC1000 is a thin ($\approx 1{\cdot}35$ mm) void-filled polyurethane polymer sheet. The voids, which do not usually appear to be interconnected, average 30 microns in diameter and occupy about 35% of the volume of the material (Figure 2). When cleanly sliced along a plane, the surface of the pad is about 50% covered with exposed pores and is naturally rough. Conditioning adds to this natural roughness in the interpore areas. Evidence of the latter can be seen in the background at the top of Figure 2. The interpore

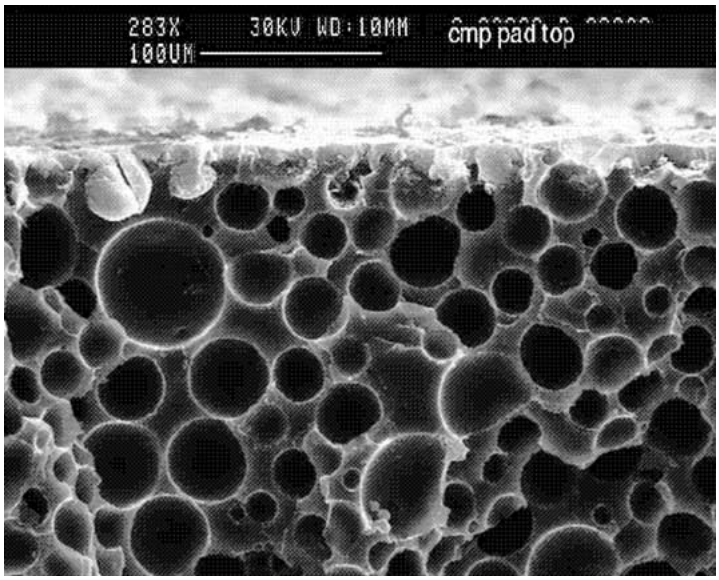

Figure 2. Scanning Electron Micrograph (SEM) cross-section of a used, conditioned IC1000 polishing pad. Surface asperities can be seen in the background at the top of the image. (Data by Letitia Malina, Motorola)

asperities, or surface summits, are the pad features that actually make contact with the surface of the wafer. We note that conditioning also produces gradual thinning and shaping of the pad surface. While shaping can cause center-to-edge polish-rate variations on a wafer, it has little impact on the average polish rate.

The wafer and the surface materials that are usually polished have elastic moduli that are on the order of 100 GPa. By comparison, the IC1000 pad is very soft. Figure 3 shows the modulus of a dry IC1000 sample as a function of temperature and time as measured with a dual cantelever constant-strain test on a dynamic mechanical analyzer (DMA). In this test, a fixed strain is applied to a pad sample and the reaction force from the pad is measured as a function of time. It can be seen that the IC1000 behaves viscoelastically since the reaction force decays with time. Relaxation to a point half way between the initial modulus and the equilibrium modulus occurs in about 2–4 minutes. The pad modulus is also sensitive to temperature. Furthermore, water is a plasticizer for polyurethane, so it possible for the relaxation modulus of a well-soaked pad to be less than that of a dry pad by a factor of ≈ 3–5.

In a polishing experiment, the relative sliding velocity between the pad and wafer is around 1 m/sec. For an 8-inch-diameter wafer, the actual contact time between a fixed point on the pad and the wafer surface is consequently no more than about 0·2 sec, which is much less than the time scale for viscoelastic relaxation. Thus, we will treat the pad as if it were elastic and will ignore residual strain that may accumulate in an asperity on successive passes under the head. We will assume that the pad surface is being washed with room-temperature slurry and that its temperature is 20 C.

3. Pad/wafer contact model

Greenwood and Williamson [3] developed a model for contact between a smooth flat surface and a rough surface using Hertzian contact theory [4, Chapters 4, 11]. In the Greenwood and Williamson model, the surface of the pad consists of a population of summits, or asperities,

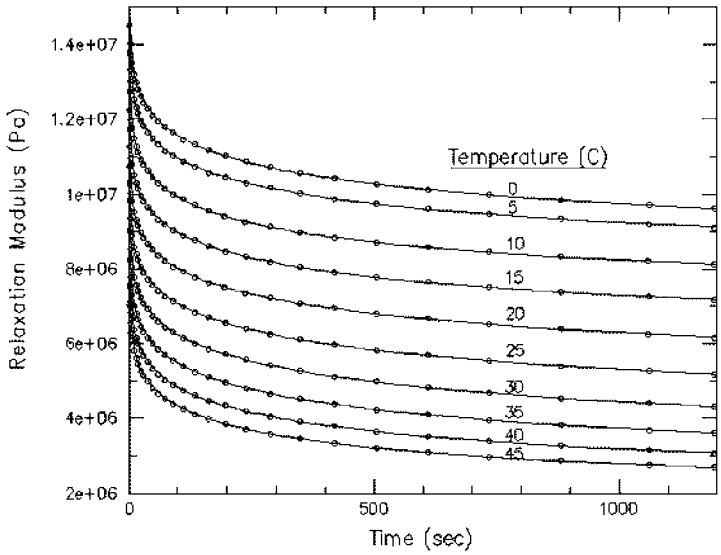

Figure 3. Dynamic Mechanical Analyzer (DMA) raw relaxation-modulus data for the IC1000 as a function of temperature. The actual modulus can be obtained by multiplying each curve by 50. (Data courtesy Diane White, Motorola)

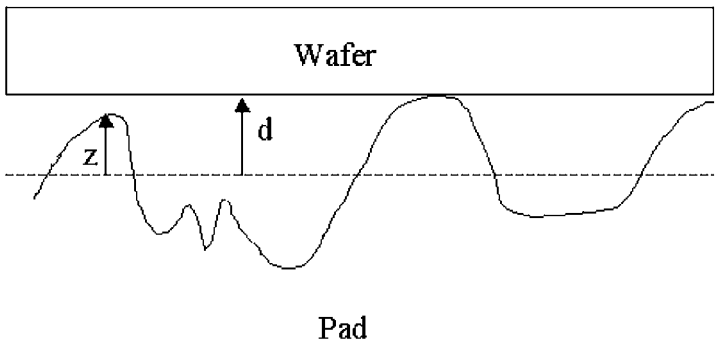

Figure 4. A flat wafer in contact with a rough pad surface. The separation d between the pad and wafer is measured from the mean plane of the pad surface. Asperity heights are also measured from the mean plane.

of area density η_s whose heights relative to the mean plane of the surface have probability-density function (pdf) $\phi(z)$ (Figure 4). All asperities are assumed to have spherical tips of identical curvature κ_s and are presumed to behave elastically. If an asperity has height z and the distance of approach between the wafer and the mean plane of the pad is $d < z$, then in the Hertz theory it will carry the load

$$L = 4/3E^*/\kappa_s^{1/2}(z-d)^{3/2} \tag{1}$$

over a circular contact region with area

$$A = \pi(z-d)/\kappa_s, \tag{2}$$

where $E^* = E/(1-\nu^2)$, E is the pad Young's modulus and ν ($\approx 0{\cdot}5$ for polyurethane) is the Poisson ratio.

If there are N summits in a nominal area of A_0, so that $\eta_s = N/A_0$, then it follows that the nominal pressure (the total load divided by the nominal area) at separation d is

$$\bar{p} = 4/3\eta_s E^*/\kappa_s^{1/2} \int_d^\infty (z-d)^{3/2}\phi(z)\mathrm{d}z \tag{3}$$

and the ratio of the actual contact area to the nominal area is

$$A_f = \pi\eta_s/\kappa_s \int_d^\infty (z-d)\phi(z)\mathrm{d}z. \tag{4}$$

The above two equations form the Greenwood-and-Williamson model.

Asperity abrasion is the critical element in the current-rate decay model, so the pdf ϕ in the above model depends on time; *i.e.*, $\phi = \phi(z, t)$. Polishing experiments are also generally performed at a constant nominal pressure $\bar{p}$. The separation $d(t)$ required to maintain load balance in Equation (3) is therefore time-dependent. It is also likely that the curvatures of summits in contact with the wafer decrease over time at a rate that depends on their initial height, so that κ_s should in fact be a function of z and t inside of the above integrals. Worn summit tips will in general be non-spherical, complicating the contact analysis based on the Hertz theory. However, we find reasonable agreement with experiment without including curvature evolution.

The distinction between the nominal area of contact A_0 and the real area of contact $A = A_f A_0$ is important. The real area of contact between rough surfaces is usually very small. Optical measurements of the contact between a dye-covered IC1000 pad and a sapphire disk under loads that are consistent with CMP practice indicate that 1% or less of the pad surface is in contact with the wafer [5, pp. 81–85]. Therefore, when slurry-fluid-pressure effects are small, the actual contact pressure between pad asperities and the wafer is larger by a factor of 100 or more than the nominal contact pressure. We will connect this actual contact pressure at an asperity tip with the asperity wear rate.

4. Wear of the pad and the wafer

The formulation of an evolution equation for the asperity-height distribution $\phi(z, t)$ is straightforward. Consider an asperity of height z with constant cross setional area (appropriate if the amount worn is small relative to the height). From Archard's law [6], the rate of wear of the asperity is proportional to the actual contact pressure,

$$\frac{\mathrm{d}z}{\mathrm{d}t} = -c_a(L/A) \tag{5}$$

$$= -c_a \frac{4}{3\pi} E^* \kappa_s^{1/2} (z-d)^{1/2}, \tag{6}$$

where c_a is a parameter that is proportional to the sliding velocity of the asperity tip relative to the wafer.

The rate of change of the fraction of summits in the range $(z, z + \Delta z)$ equals the rate at which eroding summits enter the interval from above minus the rate at which they leave from below. Combining this with the above equation, we have

$$\frac{\partial}{\partial t} \int_z^{z+\Delta z} \phi(u, t)\mathrm{d}u = \frac{c_a 4 E^* \kappa_s^{1/2}}{3\pi} ((z + \Delta z - d)^{1/2}\phi(z + \Delta z, t) - (z-d)^{1/2}\phi(z, t)). \tag{7}$$

It follows by approximating the left side by $\partial/\partial t\phi(z,t)\Delta z$, dividing by Δz and taking the limit as $\Delta z \to 0$ that for $z > d$,

$$\frac{\partial \phi(z,t)}{\partial t} = \frac{c_a 4E^* \kappa_s^{1/2}}{3\pi} \frac{\partial}{\partial z}((z-d)^{1/2}\phi(z,t)), \tag{8}$$

while for $z \leq d$ the rate of change of $\phi(z,t)$ is zero. This is the evolution equation for the asperity-height distribution. It requires an initial condition, $\phi(z,0) = \phi_0(z)$.

The behavior of the solution of the above equation at large times can be understood by considering the physical picture of asperity wear at constant separation d. Initially, all of the asperities of height $z > d$ are in contact with the wafer and wear at rates that increase with height. At long times, the heights of the contacting asperities all approach d. Thus, the area under the pdf to the right of d is converted into a delta function at d while the rest of the distribution is unchanged.

Wear of the wafer is usually modeled with Preston's equation [7], an empirical relation that says that the average wear rate is proportional to the product of the nominal pressure and the sliding velocity V. Recognizing that only a fraction of the pad area is in real contact and that this fraction increases gradually as asperities wear, we instead hypothesize that the average removal rate $RR(t)$ is proportional to average real contact pressure, or equivalently that

$$\mathrm{RR}(t) = \frac{c_w \bar{p}}{A_f(t)} \tag{9}$$

where $\bar{p}$ is the nominal pressure and c_w, like c_a, is proportional to the relative sliding velocity. Like Equation (5), the above rate expression is a form of Archard's law.

There is a difference between wear of the wafer and wear of the pad asperities in that the wafer is always in contact with the pad but a given pad asperity is not always in contact with the wafer. In a CMP tool, as in Figure 1, pad-asperity wear should depend on the accumulated contact time and therefore will change with distance from the center of the pad. For simplicity, we model the wear rates at the wafer center, where the asperity-contact time is related to the wafer-contact time by a constant scaling factor.

5. Numerical method and comparison with experiment

The mathematical model describing the CMP removal rate with rate decay is then

$$\bar{p} = 4/3\eta_s E^*/\kappa_s^{1/2} \int_{d(t)}^{\infty} (z-d(t))^{3/2}\phi(z,t)\mathrm{d}z, \tag{10a}$$

$$A_f(t) = \pi\eta_s/\kappa_s \int_{d(t)}^{\infty} (z-d(t))\phi(z,t)\mathrm{d}z, \tag{10b}$$

$$\frac{\partial \phi(z,t)}{\partial t} = \frac{c_a 4E^* \kappa_s^{1/2}}{3\pi} \frac{\partial}{\partial z}((z-d(t))^{1/2}\phi(z,t)), \tag{10c}$$

$$\mathrm{RR}(t) = \frac{c_w \bar{p}}{A_f(t)}. \tag{10d}$$

This set of equations can be solved numerically with the following procedure:
1. Initialize $\phi(z, t)$ to $\phi_0(z)$;
2. Solve for the load-balancing separation $d(t)$ using (10a);
3. Calculate the actual contact-area fraction $A_f(t)$;
4. Integrate the evolution equation for the pdf for one time step;
5. Calculate the removal rate;
6. Increment the time and repeat from (2) until done.

The data that we will use for comparison with the model is the total removed over successive time itervals. The total amount removed, $\mathrm{TR}(t_0, t_1)$, between time t_0 and t_1 can trivially be obtained by integrating the instantaneous removal rate,

$$\mathrm{TR}(t_0, t_1) = \int_{t_0}^{t_1} \mathrm{RR}(t)\mathrm{d}t. \tag{11}$$

The main difficulty in the procedure outlined above is in the integration of Equation (10c), which is similar to a Hamilton-Jacobi equation. Even at very small time steps, explicit numerical schemes eventually develop instabilities as the evolving tail of the pdf approaches a delta function. However, we were able to obtain reasonable solutions over most of the duration of an experiment reported in [1] using standard fourth-order Runge-Kutta for time integration and centered finite differences for the spatial derivative.

In [1], silicon-dioxide-coated wafers were polished for a total of 60 minutes without pad conditioning. The pad surface was characterized initially and at five-minute intervals of polishing using automated optical interferometry. The surface-characterization data included the standard deviation of surface heights, the skewness and the kurtosis. This is sufficient for description of the surface pdf using a Pearson IV distribution. A plot of the initial surface-height pdf constructed from this data is shown in Figure 5. In the absence of asperity-height data, we use this function for ϕ. The right-hand tail of the pdf is the portion in contact with the wafer and the mean is located at the mean plane of the pad surface. The shape of the left-hand tail is determined mainly by the pad-pore structure while that of the right-hand tail is influenced by conditioning, wear and possibly by plastic deformation.

To supply the remaining parameters that are needed for the model, in the absence of data from Stein we take values of the asperity density $\eta_s = 2 \times 10^8/\mathrm{m}^2$ and summit curvature $\kappa_s = 2 \times 10^4/\mathrm{m}$ from [5]. These parameters are difficult to measure, vary considerably from pad to pad and are dependent on the details of the conditioning process. We used for the room-temperature pad modulus $E^* = 119$ MPa, which has been been reduced from DMA data by a factor of 5 to account for soaking in water. For the nominal contact pressure, we use Stein's $\bar{p} = 50$ kPa. The wear coefficients c_a and c_w are treated as adjustable parameters. They were selected so that the modeled amount removed agrees with the measured amount at 5 and 10 minutes. This occurs when $c_w \approx 1{\cdot}6 \times 10^{-16}$m/sec/Pa and $c_a \approx 2{\cdot}5 \times 10^{-16}$m/sec/Pa. These values are consistent with the plausible notion that the asperities wear faster than the chemically attacked wafer surface.

Using the selected parameters and pdf, the calculated total thickness removed at five-minute intervals is compared in Figure 6 with the measurements of [1]. At 45 minutes, further integration of Equation (10b) with fourth-order Runge-Kutta becomes too unstable to continue. It can be seen that the solution up to this point reproduces the time dependence of the decline in the amount removed beyond the first two data points, which were used to estimate the wear coefficients. Since the removal rate is proportional to the reciprocal of the contact

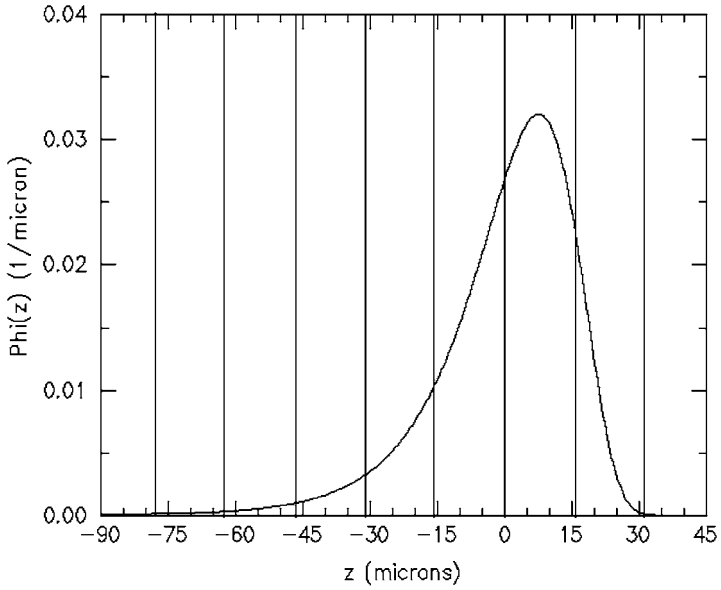

Figure 5. Pdf of surface heights constructed from data in Stein [1]. The mean of the pdf is 0, and the standard deviation, skewness and kurtosis (taken as in Stein to be the fourth moment over σ^4) are 14·6 microns, −0·98 and 4·1, respectively. The vertical lines divide the graph into $\approx 1\sigma$ strips.

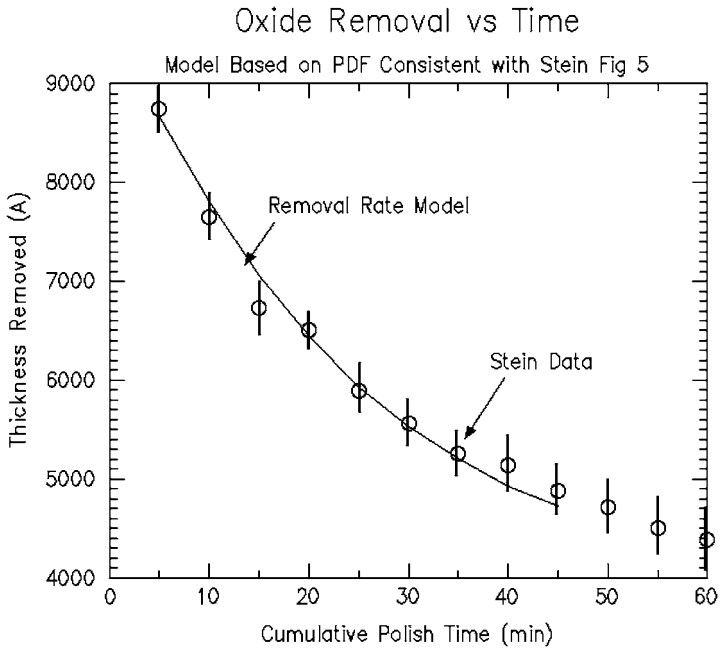

Figure 6. Amounts of silicon dioxide removed calculated using the present removal-rate model compared with data from Stein.

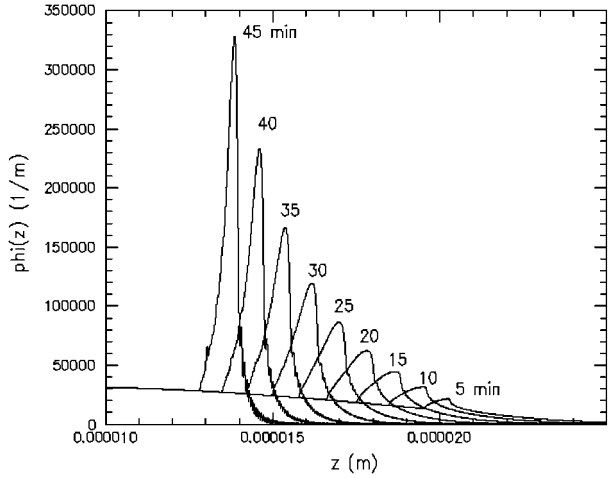

Figure 7. Evolution of the tail of the pdf in Figure 5 at five minute intervals using Equation (10c). Developing instabilities are evident in the pdf tails at longer times.

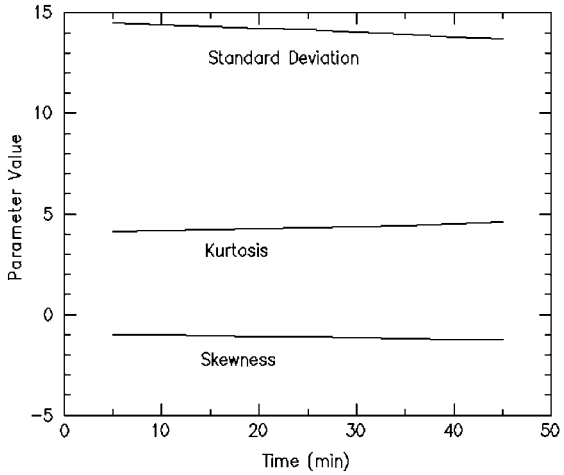

Figure 8. Calculated standard deviation, skewness and kurtosis of the distributions in Figure 7. The standard deviation is in microns; the skewness are kurtosis are unitless.

area fraction, rate decay in the current model can be traced to an increase in the contact area. In this instance, the contact area fraction is predicted to increase from 0·29% at five minutes to 0·51% at 45 minutes. These magnitudes are consistent with the data given in [5].

Figure 7 shows the evolution of the tail of the original pdf at five-minute intervals. It can be seen that the secondary peak that occurs due to wear of tall contacting asperities moves to the left and becomes progressively narrower and higher as time passes. The location of the peak is approximately at the load-balancing separation d at each time. Thus, the separation between the pad mean plane and the wafer decreases as the taller asperities wear away. The total decrease in d over the 45 minute simulation is 7·9 microns. In [2], a decrease in interpore near-surface asperity height of 1·5 microns was observed after eight minutes of polishing (with a different slurry) at twice the plate speed and about half of the applied load. Preston's equation implies that the removal rates in the Oliver and Stein experiments should be comparable. In fact, we find a 1·8 micron decrease in d after eight minutes *vs* Oliver's 1·5 microns.

Finally, Figure 8 shows the evolution of the standard deviation, skewness and kurtosis of the distributions in Figure 7. It can be seen that these parameters are quite stable with time. The calculated variations are well within the measurement accuracy of [1].

6. Summary and conclusions

We have presented a model for decay of the average polish rate in chemical-mechanical polishing in the absence of pad conditioning. The model is based on Greenwood-and-Williamson contact mechanics and on Archard's wear law. The effect of wear on the polish rate occurs though an increase in the real contact area fraction with time. The contact area in turn depends on the asperity-height pdf, which evolves according to a Hamilton-Jacobi-like equation. Using experimentally realistic values for the initial pad-roughness statistics, the pad elastic modulus and the asperity density and tip curvature, we have found that the model correctly yields the observed time dependence of rate decay using two adjustable rate parameters. Predictions of the contact-area fraction and the amount of wear of the highest asperities also agree with data. Finally, the model explains why changes in the overall surface-roughness statics may not be observed, thus resolving the apparent conflict between the Stein-and-Oliver experiments.

Aside from explaining rate decay, the value of this model lies in the connection that it makes between the pad properties and roughness statistics and polish-rate decay. The pdf of the pad surface is determined by the pad-pore structure and by the construction and operation of the conditioning tool. Thus, it may be possible to use the model in conjunction with a conditioning model to predict how different conditioning strategies will affect polish rates. Control of polish rates and polish-rate stability is the bottom line for industry.

7. Afterword

Nearly a year after this paper was written, the traveling secondary wear-induced peaks predicted in this paper were detected using high-resolution optical interferometry [8].

References

1. D.Stein, D. Hetherington, M. Dugger and T. Stout, Optical interferometry for surface measurements of CMP pads. *J. Electr. Materials* 25 (1996) 1623–1627.
2. M. R. Oliver, R. E. Schmidt and M. Robinson, CMP Pad-surface roughness and CMP removal rate. In: R. L. Opila, C. Reidsema-Simpson, K. B. Sundaram and S. Seal (eds.), *Proc. 198th Meeting of the Electrochemical Society, 4th International Symposium on CMP*. Phoenix, Arizona, 22-27 October (2000) pp. 77–83.
3. J. A. Greenwood and J. B. P. Williamson, Contact of nominally flat surfaces. *Proc. R. Soc. London* A295 (1966) 300–319.
4. K. L. Johnson, *Contact Mechanics*. Cambridge: CUP (1985) 452 pp.
5. L. Shan, *Mechanical Interactions at the Interface of Chemical Mechanical Polishing*. PhD thesis, Georgia Institute of Technology (November 2000) 198 pp.
6. J. F. Archard, Contact and rubbing of flat surfaces. *J. Appl. Phys.* 24 (1953) 981–988.
7. F. W. Preston, The theory and design of plate glass polishing machines. *J. Soc. Glass Technol.* 11 (1927) 214–256.
8. A. S. Lawing, Polish rate, pad surface morphology and pad conditioning in oxide chemical-mechanical polishing. *Proc. of the 2002 Materials Res. Soc. Spring Meeting*. San Francisco, California, April 1–5, 2002.